BOLT BERANEK AND NEWMAN INC

CONSULTING · DEVELOPMENT · RESEARCH

AFCRL-68-0120

AD670604

ON THE PSYCHOLOGICAL IMPORTANCE OF TIME IN A TIME SHARING SYSTEM

Jaime R. Carbonell
Jerome I. Elkind
Raymond S. Nickerson

Contract No. F19628-68-C-0125
Project No. 8668
Task No. 8668-00
Work Unit No. 8668-00-01

Scientific Report No. 6

14 September 1967

This research was sponsored by the Advanced Research Projects Agency under ARPA Order No. 627.

Contract Monitor: Hans Zschirnt
Data Sciences Laboratory

Prepared for:

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

30

AFCRL-68-0120

ON THE PSYCHOLOGICAL IMPORTANCE OF TIME IN A TIME SHARING SYSTEM

Jaime R. Carbonell
Jerome I. Elkind
Raymond S. Nickerson

Contract No. F19628-68-C-0125
Project No. 8668
Task No. 8668-00
Work Unit No. 8668-00-01

Scientific Report No. 6

14 September 1967

This research was sponsored by the Advanced Research Projects Agency under ARPA Order No. 627.

Contract Monitor: Hans Zschirnt
Data Sciences Laboratory

Prepared for:

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

ABSTRACT

One of the most important problems in the design and/or operation of a computer utility is to obtain dynamical characteristics that are acceptable and convenient to the on-line user. In this paper we are concerned with the problems of access to the computer utility, response time and its effect upon conversational use of the computer, and the effects of the load on the system (and its fluctuations) upon the other aspects.

Primary attention is placed upon response time. Some of the difficulties in its definition are pointed out through examination of the typical interaction process. It is concluded that rather than a single measure a set of response times should be measured in a given computer utility, in correspondence to the different types of operations requested. Next, it is tentatively assumed that the psychological value of short response times stems from a subjective cost measure of the user's own time, largely influenced by the value of concurrent tasks being postponed. A measure of cost (to the individual and/or his organization) of the time-on-line required to perform a task might thus be derived.

More subtle is the problem of the user's acceptability of given response times. This acceptability is a function of the service requested (e.g., length of computation), and variability with respect to expectations due both to uncertainty in the user's estimation and to variations in the response time originated by variable loads on the system.

This paper presents a strong advocation that an effort be made by computer-utility designers to include dynamic characteristics (such as prediction of loads and their effects) among their design specifications. To achieve this goal, more research both on the human factors and systems aspects of this problem is urgently needed.

Among the most important problems facing designers of computer utilities is that of determining the relationship between the dynamical characteristics of the system and its acceptability to the user. In this paper we will discuss two dynamical aspects of time sharing systems, namely: accessibility and response time.

It is obvious that users would always like very rapid response times and continuous accessibility to the service; however, technological and economical constraints often conflict with both of these desires. In actual installations we may be faced with dynamical characteristics that range from highly satisfactory to completely unacceptable. Though it is relatively simple to state what ideal conditions are, degrees of acceptability are elusive and complex to characterize. In particular it is very difficult to draw the line between that which is just tolerable and that which is not. For this reason, we do not aspire in this paper to provide any definite answers, but only to identify some of the problems, to show their implications and importance and to suggest some possible ways of dealing with them.

At the present time, designers and managers of time-sharing computer utilities are forced to discover after the fact, through the reactions of users, the acceptability of the dynamical characteristics of their systems. Hopefully thorough investigations of these characteristics, both from a systems and a psychological point of view, may lead to the establishment of performance criteria which could help guide system development from the early phases of design.

Let us consider first the problem of accessibility, which is, of course, a function of the characteristics of the load on the time sharing system. The demand for service depends on the time of the day. A recent study[1] on the Bolt Beranek and Newman commercial time-sharing system (Telcomp) yielded the starts by hour of the day, presented in Figure 1. The population of users generally belongs to the scientific and engineering communities. The influence of regular working hours, and the low corresponding to lunch breaks, are apparent. The length of a console session was also measured for the different starting hours. As was to be expected, longer sessions correspond on the average to early morning, early afternoon, and late evening starts, the shortest ones to late morning and especially late afternoon starts. The frequency of the different durations (regardless of hour of day) can be plotted against the duration as has been done utilizing semilogarithmic coordinates in Figure 2. The good approximation by a straight line indicates the quasi-exponential character of the functional dependence; a similar result was found in other scientific time sharing systems such as CTSS at MIT. This seems then to be an approximate characteristic of many known users' populations. An even more interesting result is obtained if the mean duration of a console session is computed. A mean duration around one hour is obtained for the console session in most systems with known statistics in scientific applications. The Telcomp system gives about one hour and ten minutes; incomplete results on the new BBN time-shared research computer, with a much greater percentage of computer-oriented users, still indicate an average console session in the same range, though particular types of

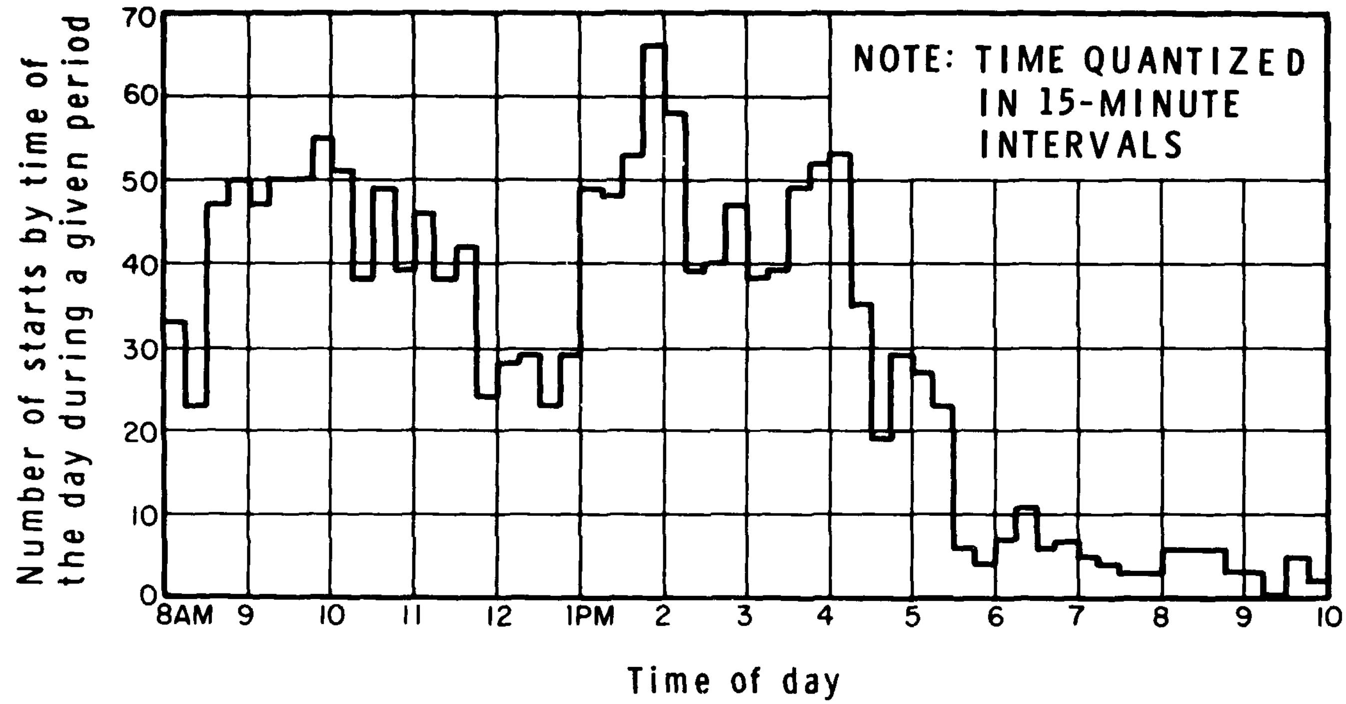


FIG. 1 DISTRIBUTION OF STARTS BY TIME OF THE DAY IN A TIME SHARING SYSTEM (TELCOMP, OPERATED BY BOLT BERANEK AND NEWMAN INC.)

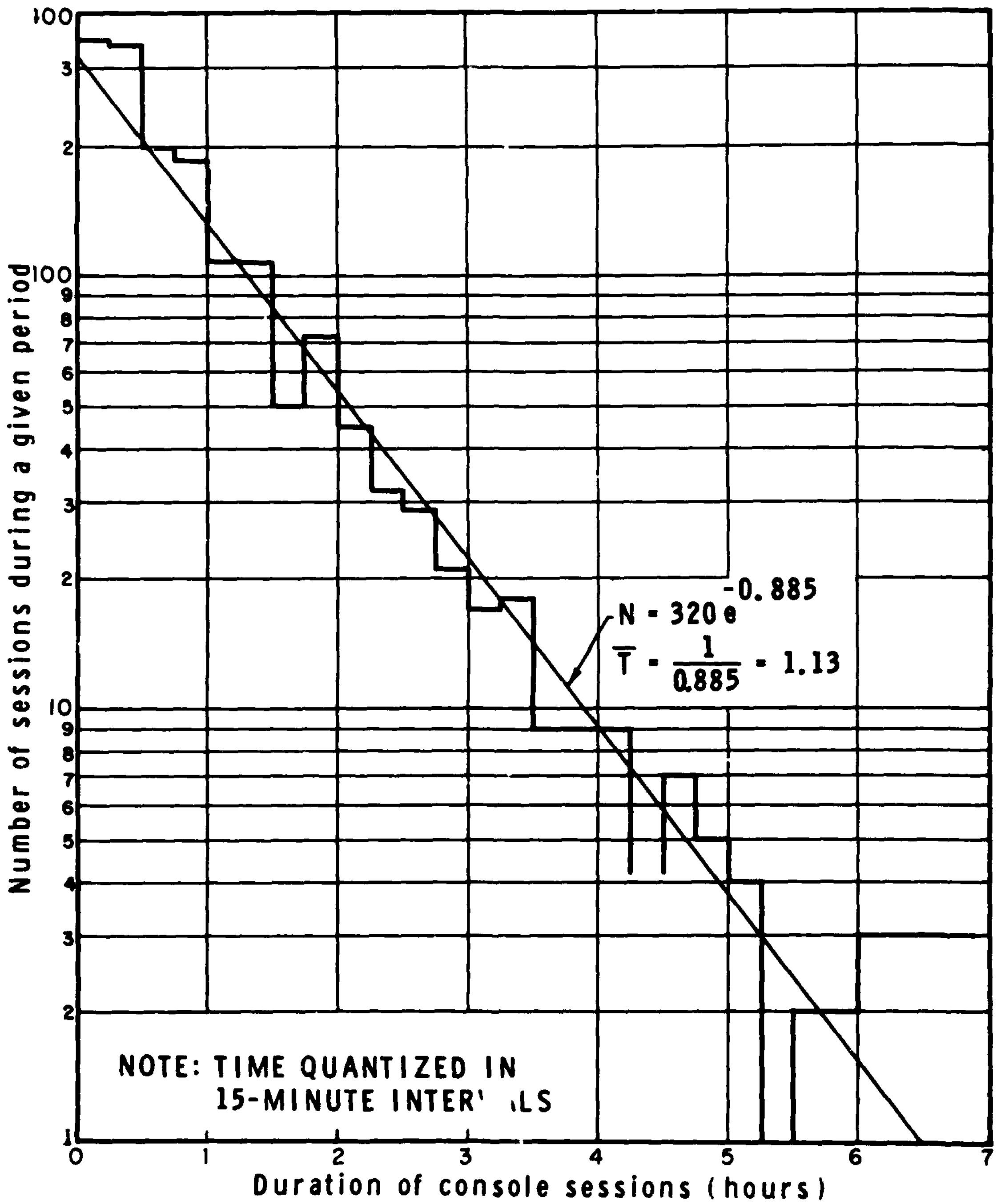


FIG. 2 DISTRIBUTION OF DURATION OF CONSOLE SESSIONS IN A TIME SHARING SYSTEM (TELCOMP, OPERATED BY BOLT BERANEK AND NEWMAN INC.)

users yield values well above or below the average. The approximate constancy of the general shape and mean of the distribution of work session durations suggests the influence of some basic psychological or even physiological constants on users' behavior.

As in the case of co-workers in a project, the man should have easy and continuous access to his partner, the computer, if the state of man-computer symbiosis[2] is to be achieved. There are several factors, however, that may tend to limit accessibility by a user to a computer system. These may be an occupied shared-terminal, restrictions placed on access by the management to avoid possible overloads on central resources (even if lines are available)[3], the more or less inevitable "down" times originated by malfunctions and system maintenance and improvements, or a high overall demand on the system occupying all communication lines.

Let us consider what happens, for example, in the latter case. A user dials in and gets a busy signal, i.e., is informed that he must wait to get on-line. Two factors suggest themselves as being possible determinants of his degree of frustration with the system: (1) the percentage of times he gets a busy signal and (2) the expected length of time he has to wait in line (after getting a busy signal) before eventually getting access to the computer. As an example of the importance of the selection of the appropriate criteria of acceptability, let us consider the case of a study of accessibility to an early version of the Telcomp system[4]. The objective was to find limits on the acceptable load based on "reasonable" access;

in other words, because of having a limited number of lines and assuming random usage on the part of the users, we wanted to determine how many subscribers should be given access to the system. Two criteria had been tentatively proposed by the management: a 10 percent occurrence of busy signals (on the average), and an expected wait in queue of 15 minutes after receiving a busy signal. The analytical study showed, among other things, that the proposed waiting time criterion was in fact much less restrictive on the system conditions than the one based on a percentage of busy signals. Because of the averaging effect of random arrivals and departures, the expected wait was rather short even under high percentage of busy signals. This showed that one at least of the two proposed criteria should be revised or dropped; it really illustrates something more, namely the need for basic theoretical and experimental psychological studies on the development of acceptability criteria. If we had these available, which we do not, we could answer in general terms or in the context of particular systems, the following important practical questions: Which criterion is the best from a human factors point of view? Is there some psychological correlation between percentage of busy signals and expected wait after getting a busy signal? Is it valid to take expected values or should we consider the tails or shape of distributions as well? What percentage of the users' population can a time sharing system afford not to satisfy? Unfortunately, all these questions remain unanswered at this time.

Let us now consider what is our second major concern in this paper, the so-called response time of a time sharing

system. Response times are one of the most important elements influencing users' behavior, the amount of work they are able to accomplish, and their degree of satisfaction with a time-sharing system. Response times are not the only consideration, of course, since many other characteristics of a system will also have strong effects; for example, languages available on-line, command language, file system, interface characteristics, etc. We have chosen in this paper to concentrate on time-dependent aspects and their effects, and particularly, on the so-called response time.

In a discussion of response time, it may be helpful to draw an analogy between the man-computer interaction and a conversation between two humans. What characteristics must a communication process have to be called a conversation? First of all, there must be some degree of symmetry, some balance between the roles of the two participants. In this sense, reading a book, or the presentation of a paper are not conversations. Secondly, a conversation has the property of continuity, if the gaps between talk spurts get long enough we simply no longer have a conversation. In human conversations there is a quite apparent intolerance for even relatively short periods of silence. This intolerance, however, is to some extent situation dependent: if one feels he has lost the attention of his partner he gets irritated, but if he feels that his partner is busy preparing a thoughtful response he will be willing to tolerate longer delays. To alleviate this situation, the listening partner in a human conversation tries to reassure the other that he is attending by means of sporadic assertions, repetition of some key words, etc. In a conversation it is most frustrating to feel that

our partner is not giving us his full attention because he is time sharing himself: signing letters, calling his secretary for other matters, independently of his participation in the conversation. This human desire to capture the full attention of the other side in a conversation brings the question of how important it is from the psychological point of view to keep the sharing of the computer attention invisible to the user in order to create the illusion that he is the only one being served. If delays in responding are too long this illusion is not created.

When we refer to the man-computer interaction in the time shared environment as being a conversational-type interaction, we are probably alluding as much as anything else to the time scale on which the interaction takes place. This is the time scale that is characteristic of interpersonal conversations in contrast with that of a postal correspondence. There is, of course, a two-way exchange of information in the ease of postal correspondence and similarly, there is a dialogue of sorts even in the conventional batch processing computer operation. However, in the conversational-mode operation of a time shared system, the "response time" of the system is orders of magnitude shorter. But what exactly do we mean by the response time of time sharing system?
Here is a case in point of a need for clarifying our descriptive terminology. The term has been used in a rather loose way in the past, the imprecision being utilized for their own benefit by those making unjustified claims about their systems. The response time is usually assumed to be the time elapsed from entering a command until its completion, the latter being characterized by the production of an output or

other signal to the user, and the transfer of control to him; there may be some output bursts during command execution which, contrary to the usual notion of response time in Psychology, are included in the response period without necessarily implying termination. The term response time sometimes refers to a particular occurrence of a particular computational request, while in other cases it refers to an aggregate measure on a time sharing system in general. In the latter case, the response time should not be measured by a unique figure, but rather by a set of values. Response times depend on a number of factors, which should be mentioned when the values are presented, and furthermore, should be standardized if one wants to compare different systems. First we have the amount of computation requested by the user, as well as the system components and resources involved. The exact output requested and the instant in its production used to terminate the time being measured also constitute important factors. Finally there is the most notable one, namely the effect of variable loads on the computer. In evaluating the response characteristics of a time sharing system, different conditions must be tested, including at least response to a token command (very short computations), to a long standardized computation, and to file manipulations involving two-way access to the different types of secondary storage available in the utility; these requests should be tested under different system load sizes and configurations.

How can we evaluate the degree of acceptability of response times? Let us start with an informal discussion to illustrate some of the main effects.

For simple requests the acceptable delay is probably a constant of the order of a few milliseconds, similar to delays encountered in fast conversational interchanges between people in which case, the computer will look no worse than the human counterpart. For requests involving long computations, the tolerance for delays will depend on the user's expected net (unshared) computation time vs. the actual delay, as well as on the rate at which useful information comes out as an output from the computer. The latter will help occupy the user's mind and allow him to do useful work while waiting. The expedience of input/output facilities is also of great importance, and part of the benefits from graphical input/output devices are related to temporal effects.

Long waits defeat one of the basic reasons for time-sharing, that of matching human and computer speeds; if waits are too long, the user will probably choose to time-share himself by switching to another task (like reading a book, or even controlling a second time sharing terminal) during the delays. The tolerance for a given response time will be lower the busier the user is in his professional life, as a reflection that the cost of his own time depends on other tasks being postponed by the current one. More on this later.

Users of a time sharing system particularly dislike unpredictable response times (due to variable loads on the system). It has been observed that they usually prefer a constant delay to a possibly shorter but variable one; unpredictable conditions disturb the user and interface with his efficient use of the computer. The above assertion can be interpreted by saying that if delays are long but predictable,

a user can conceivably carry on some other activity instead of wasting time waiting for a result that may come now or later; this is closely related to the concept of human memory swapping, as presented by Simon[5]. Furthermore, if the mean and variance of the distribution of response times under a given condition are known, the acceptability is probably higher than when they are unknown.

The informal considerations presented above can help us plan an attack on the problem of degree of acceptability of response times. From an experimental point of view, we would like to identify some measure directly related to acceptability. We also need to have some basic model to guide our efforts in the design of experiments. One conceptual framework could be provided by the notion of total work accomplished with respect to a variety of tasks (one or more of them on line) that are competing for man's attention. Assume that the human operator weights the different tasks and thus derives what we may call a set of "costs" of not performing them. One can then think of experiments in which an on-line, well-defined, measurable task competes with some other measurable task for the individual's attention. Measuring the work accomplished under controlled conditions (including controlled variations in the time sharing system characteristics with respect to response-time and probably other aspects) provides an objective estimation related to acceptability. Such an objective measure should be complemented by a subjective measure of acceptability on some scale by interrogating the subjects about their degree of satisfaction with the system. Subjective estimations will provide a checking mechanism on how representative of acceptability the objective measure is.

It will be useful at this stage, to treat in a more formal way the work accomplishment and the degree of satisfaction of a user with respect to a time sharing system by attempting to identify some of the factors affecting them. These are numerous and of widely different kinds. We shall limit ourselves, however, to those related to response time. Within these we shall attempt to present some which, from our point of view, and according to our previous informal discussion, are of prime importance.

Without loss of generality, the degree of acceptability, say Θ, of response times can be expressed as a function of a number n of parameters, each corresponding to one of those factors. As shown in Figure 3, where we have taken n=3 for illustrative reasons, we can define an n-dimensional space where each set of characteristics (or parameter values) determines a point with a corresponding value of acceptability Θ . The locus of the points of given equal acceptability is a (hyper) surface in the n-dimensional space. Points P and Q, both on the Θ=A surface, illustrate the effect of a trade-off between different factors; in other words, the same degree of acceptability may be obtained for different sets of parameter values. The trade-off between the variability of reponse times and the corresponding expected value is an important example of this effect.

In talking of subjective acceptability we must make a distinction between different forms of it. First we could talk of $\Theta_{k,u,T}$ corresponding to a given computer task k, to a given user u, and to a given instant of time T. Next we have $\Theta_{k,u}$ which characterizes the acceptability for the same user u and

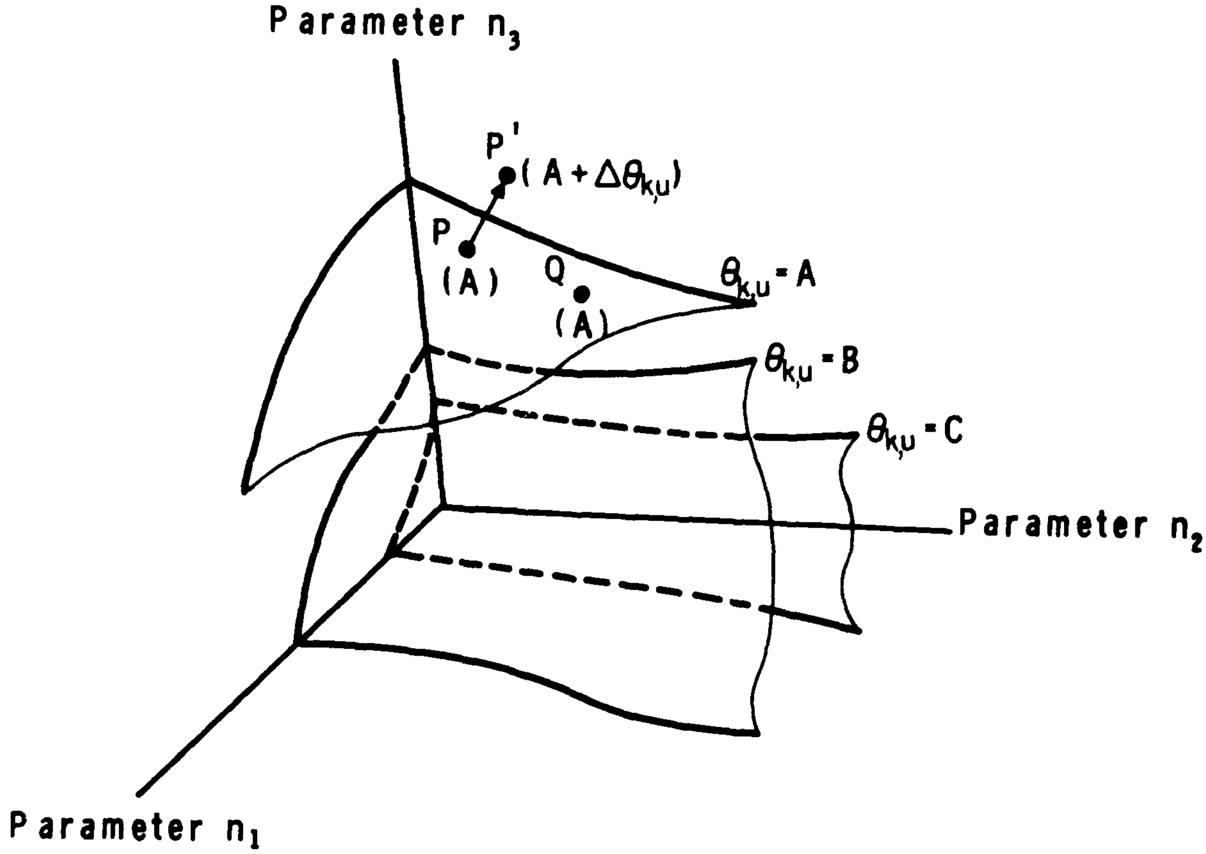


FIG. 3 DEGREE OF ACCEPTABILITY AS A FUNCTION OF THREE HYPOTHETICAL PARAMETERS

the same task k, but without specifying a particular occurrence, so it may be the averaged result of many repetitions of the same computation. As a third possibility we have θ_u which averages the acceptability of a system by a given user across tasks and repetitions. Finally, we could talk of θ which corresponds to the degree of acceptability (in terms of response times) of a given time sharing system in general, for a population of users.

Similarly to the θ's we could talk of $W_{k,u,T}$, $W_{k,u}$, W_u and W for the work accomplishment. In experimental work we would want to measure in each session sets of $W_{k,u,T}$'s and ask for sets of $\theta_{k,u,T}$'s, which would give in that experimental session a composite datum point for $W_{k,u}$ and another for $\theta_{k,u}$. The correlation between W's and θ's is an important point to investigate.

To decide which temporal parameters have a prime effect on work accomplished and/or degree of acceptability during a console session is a difficult task. Based on observations and discussions, including our personal experience as users of several time sharing systems, we have tentatively identified some of those factors. Without any pretension of exhaustiveness or hierarchical ordering, our list includes: central tendency and dispersion measures of observed response times; central tendencies and dispersion measures of times required for input and of times devoted to the production of useful, informative output; some parameters related to the importance of the on-line task with respect to the user's overall occupations and time availability.

With these factors in mind, we can write the following tentative functional relationships for the acceptability $\Theta_{k,u}$ and the work accomplished $W_{k,u}$ for a given user and a given application:

$$\Theta_{k,u} = F_{k,u} (T_r, t_r, T_o, t_o, T_i, t_i, T_a, T_t, \ldots..) \qquad (1)$$

$$W_{k,u} = G_{k,u} (T_r, t_r, T_o, t_o, T_i, t_i, T_a, T_t, \ldots..) \qquad (2)$$

In these equations the sets of periods represent possible dependences from other parameters apart from the proposed ones. Subindices correspond to the different factors. Capital letters and lower case ones respectively represent central tendency measures (for instance means, medians, etc.) and dispersion measures (for instance, standard deviations) in the probability distributions of values of the parameters over repetitive conditions across time. Subindex r corresponds to observed response times, so T_r is the central tendencey and t_r a measure of dispersion of those times (the dispersion being originated because of variable loads on the system). Similarly, subindex o corresponds to useful, informative output (if the output is irrelevant or highly redundant, its effect may be the opposite), and subindex i to input times. T_t is the total time devoted by the user to his on-line task, while T_a is the total availability of his time to perform tasks either on-line or otherwise. In the experiment suggested above, T_t represents the time on the teletype, while the difference $T_a - T_t$ represents the time devoted to a "secondary" task; through suitable instructions and feedback to the subject, the latter can be made either less or more important than the on-line "main" one. Thus far we have not provided any insight

into the form of the functionals $F_{k,u}$ and $G_{k,u}$ of Eqs. (1) and (2). Neither have we shown precisely how to relate them to feasible experimentation. From an experimental point of view, it is reasonable to start with a given time sharing system with some set of characteristics determining an operating point like P in Figure 3. It is feasible and useful then to design controlled experiments in which we only change some selected parameters at a time, and probably not in a drastic way. We are thus led to examine incremental variations in the work output $W_{k,u}$ and in the acceptability $\theta_{k,u}$ rather than absolute values of them. This involves measurements and judgments that are simpler than absolute ones. The total variation in a number of these experiments leads us in Figure 3 from operating point P with acceptability $\theta_{k,u} = A$ to point P' with acceptability $\theta_{k,u} = A + \theta_{k,u}$. Under suitable differentiability assumptions and for small increments, we can use Taylor's expansion to approximately express the incremental variations $\Delta\theta_{k,u}$ in the acceptability measure and $\Delta W_{k,u}$ in the work accomplished as:

$$\Delta\theta_{k,u} = \sum_{p_j} \frac{\partial F_{k,u}}{\partial p_j} \Delta p_j \tag{3}$$

$$\Delta W_{k,u} = \sum_{p_j} \frac{\partial G_{k,u}}{\partial p_j} \Delta p_j \tag{4}$$

where p_j stands for the different parameters in Eq. (1). The experiment suggested above, letting one parameter change at a time, amounts to the determination of the corresponding partial derivatives in Eqs. (3) and (4), i.e., the coefficients of the parameter variations evaluated at the operating point defined by all the parameters. In general, then, the coefficients will depend on some of the other variables, but we are now in a better position to make measurements on these coefficients, and to provide some reasonable assumptions about some of them. Eqs. (3) and (4) indicate the effects of trade-offs. In this sense, for instance, we can ask the following questions: What incremental variation t_r (say, a decrease) in the dispersion t_r of the response time will offset an incremental variation T_r (increase) in the expected response time T_r? The answer to this fundamental question, and others equally important, resides only in careful and extended experimentation.

We can next attempt to discuss briefly some of the coefficients in Eqs. (3) and (4), though we are conscious of the risk of being grossly inaccurate at this stage. For example, by following our notion of balance between competing tasks, we can talk of the "cost" of working on line in relation to other tasks being postponed, and show later how this might be related to the coefficients of T_r in Eq. (3). That "cost" C_{HO} of the user's own on-line time can be tentatively expressed as:

$$C_{HO} = \frac{a\ T_t}{1 + b\ (T_a - T_t)}\ ,\ 0 \leq T_t \leq T_a \tag{5}$$

where C_{HO} stands for human operator cost, a and b are positive constants dependent on the individual (and probably on calendar time), T_a is his total time in a suitable working period (day, week), and T_t is the total time devoted on-line to his problem as before. A plot of Eq. (5) is presented in Figure 4a.

To justify Eq. (5) on psychological grounds, let us first observe that for very small values of T_t, C_{HO} is approximately proportional to it, as it should be. The constant of proportionality is related to the cost of postponing other tasks; it will obviously be different for different users, and will also depend on the workload the particular user has for that day or week. The second term in the denominator expresses the dependence of the cost on the total available time T_a. That term is minimized (and the derivative of the total cost maximized) for $T_t = T_a$, in relation to the fact that as more and more of the available time is being taken by the on-line task, more and more important tasks are being postponed. The "1" in the denominator assures that the cost remains finite in the defined interval; in other words, 1/b may be thought of as a damping term.

Constants a and b depend on the individual. The value a is the unitary cost when all the available time is devoted to the on-line task ($T_t = T_a$), and is obviously larger the more important the man is. For a very busy scientist or executive engaged in many different and pressing tasks, b will also be quite large, since the cost of absorbing all his time may be very high (he may be a quasi-irreplaceable man). For a programmer whose task is unique and therefore cannot use his time profitably in anything else, the constant b can be quite small since there is no benefit - at least for the organization - in replacing his

waiting time with some other assignment; in this case the influence of the second term of the denominator in Eq. (5) should be small, which makes C_{HO} almost linear with T_t in the interval $(0,T_a)$.

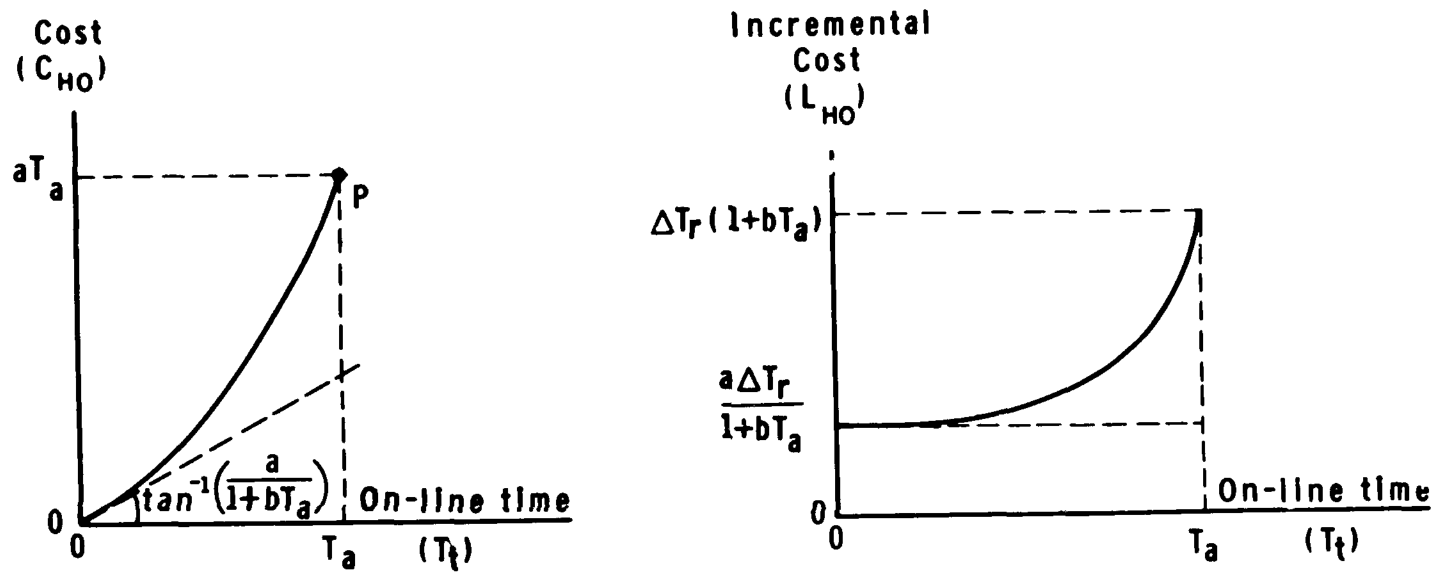
In order now to see the relationship between C_{HO} in Eq. (5) and the coefficient of T_r in Eq. (3), we can take the derivative of C_{HO} with respect to T_t, which yields a rate of increase in the user's subjective cost of time. The incremental cost which we call L_{HO} is then equal to the term in T_r in Eq. (3) and is given by:

$$L_{HO} = \frac{a(1 + bT_a)}{[1 + b(T_a - T_t)]^2}\,\Delta T_r = \frac{\partial F_{k,u}}{\partial T_r}\,\Delta T_r \qquad (6)$$

$$0 \leq T_r << T_t \leq T_a$$

This function has been plotted in Figure 4b as a function of T_t and taking ΔT_r and T_a as constant parameters. If T_r is itself much smaller than T_t we could replace ΔT_r itself. Let us also note that if the human operator has the possibility of doing something else while waiting for the computer, we could replace T_t in Eq. (6) by $\min(T_t, T_s)$ where T_s is the human swapping time as suggested by Simon[5].

In a recent study[6], it has been suggested that on-line decisions by a user as to whether to proceed with a computation or stop and do something else are based in the (perhaps unconscious) balance of two costs: the cost K if his current uncertainty about the problem he is trying to solve is not removed, and the cost L of performing the computation. The last cost is given by the sum of a computer usage cost, and



(a) Cumulative Cost

$$C_{HO} = \frac{aT_t}{1+b(T_a-T_t)}$$

(b) Incremental Cost for Given Values of ΔT_r and T_a

$$L_{HO} = \frac{a(1+bT_a)}{[1+b(T_a-T_t)]^2}\ \Delta T_r$$

FIG. 4 HUMAN COST OF WORKING ON LINE (SEE TEXT)

the human operator cost which is precisely L_{HO}, as defined by Eq. (6) above. The computer cost has been defined as a function of the utilization of the different chargeable resources in the computer utility.

Observing Eq. (6) we see that for low percentage of on-line versus available time (that is, $T_t << T_a$) and constant time availability, the term L_{HO} is linear in T, i.e., the coefficient is a constant. This may not be a bad first assumption, both in the case of T and for most of the other variables as well. If we admit that approximation, then the analysis of the data obtained in the proposed experiments may possibly amount to a multiple regression analysis.

By way of conclusion, let us say that the points presented in this paper have the character of an exploratory discussion. We have not presented experimental evidence capable of sustaining them; we have only attempted to discuss some of our thoughts, and to suggest some of the most important parameters and hopefully observable effects. There is an important task ahead for computer scientists, system designers, and man-machine specialists; they should work together in order to identify the dynamic problems related to the computer user, characterize them by means of theoretical and experimental work, and propose new and better practical solutions, as an answer to the exciting and real challenge that time sharing systems have created.

REFERENCES

1. Licklider, J. C. R., "Man-Computer Symbiosis", I.R.E. Trans. on Human Factors in Electronics, March 1960, pp. 4-11.

2. Stern, R., Unpublished Telcomp Statistics, Bolt Beranek and Newman Inc., 1967.

3. Raynaud, T. G., "Operational Analysis of a Computation Center", Unpublished S. M. Thesis, M.I.T., Dept. of Electrical Engineering, June 1967.

4. Carbonell, J. R., "A Study of Telcomp Queues", Bolt Beranek and Newman Inc., Internal Memorandum, February 1966.

5. Simon, H. A., "Reflections on Time Sharing From a User's Point of View", Computer Science Res. Review, Carnegie Institute of Technolcgy, 1966, pp. 43-51.

6. Carbonell, J. R., "On Man-Computer Interaction: A Model and Some Related Issues", Prooeedings of the 1967 I.E.E.E. System Science and Cybernetics Conference, Boston, October 1967.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Bolt Beranek and Newman Inc. 50 Moulton Street Cambridge, Massachusetts 02138	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

ON THE PSYCHOLOGICAL IMPORTANCE OF TIME IN A TIME SHARING SYSTEM

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)
Jaime R. Carbonell
Jerome I. Elkind
Raymond S. Nickerson

6. REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
14 September 1967	22	6

8a. CONTRACT OR GRANT NO F19628-68-C-0125 ARPA Order No. 627 b. PROJECT NO, Task, Work Unit Nos. 8668-00-01 c. DoD Element 6154501R d. DoD Subelement N/A	9a. ORIGINATOR'S REPORT NUMBER(S) Scientific Report No. 6 BBN Report No. 1687 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFCRL-68-0120

10. DISTRIBUTION STATEMENT
1 - Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

11. SUPPLEMENTARY NOTES This research was sponsored by the Advanced Research Projects Agency.	12. SPONSORING MILITARY ACTIVITY Air Force Cambridge Research Laboratories (CRB) L. G. Hanscom Field, Bedford, Mass.

13. ABSTRACT

One of the most important problems in the design and/or operation of a computer utility is to obtain dynamical characteristics that are acceptable and convenient to the on-line user. In this paper we are concerned with the problems of access to the computer utility, response time and its effect upon conversational use of the computer, and the effects of the load on the system (and its fluctuations) upon the other aspects. Primary attention is placed upon response time. Some of the difficulties in its definition are pointed out through examination of the typical interaction process. It is concluded that rather than a single measure a set of response times should be measured in a given computer utility, in correspondence to the different types of operations requested. Next, it is tentatively assumed that the psychological value of short response times stems from a subjective cost measure of the user's own time, largely influenced by the value of concurrent tasks being postponed. A measure of cost (to the individual and/or his organization) of the time-on-line required to perform a task might thus be derived. More subtle is the problem of the user's acceptability is a function of the service requested (e.g., length of computation), and variability with respect to expectations due both to uncertainty in the user's estimation and to variations in the response time originated by variable loads on the system.

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Human Factors						
Computer Systems						
Time-Sharing						
Man-Computer Interaction						
Response Time						
Economic Models						
Acceptability						
Work Accomplishment						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

Unclassified
Security Classification

A-31409